

A NONLINEAR FILTER CONTROL CHART FOR DETECTING DYNAMIC CHANGES

Dong Han¹, Fugee Tsung², Yanting Li¹ and Kaibo Wang³

¹*Shanghai Jiao Tong University,*

²*Hong Kong University of Science and Technology and*

³*Tsinghua University*

Abstract: All conventional control charts can be viewed as charting the output of a linear filter applied to the process data. In this paper, we present a nonlinear filter control chart (NFC) in which the control statistic consists of a nonlinear combination of a data process. We also present a theorem on an estimation of the average run length for the NFC chart, and we theoretically compare the detection power of any two such control charts. A criterion is provided for selecting an optimal NFC chart. In particular, we discuss some special nonlinear filter control charts for detecting dynamic changes in process mean that can be viewed as extensions of the conventional CUSUM control charts.

Key words and phrases: Average run length, change point detection, dynamic changes, statistical process control.

1. Introduction

Since Dr. W. A. Shewhart of the Bell Telephone Labs introduced the statistical process control (SPC) concept and developed the first statistical control chart in the 1920s, many control chart schemes have been proposed. SPC techniques utilize statistical methods to monitor each phase of the manufacturing process so as to maintain and improve the product quality while decreasing variance. There have been many applications and much development, see Montgomery (2001). This research has focused on efficient, simpler methods for detecting process changes accurately and quickly.

In manufacturing, data collected from sensors are contaminated by noise that is commonly modeled as *i.i.d.* white noise, drifts, sinusoidal noise, or general ARIMA time series noise. Conventional control charts can be viewed as charting the output of a filter applied to process data and to reveal process status. A linear filter defines a linear relationship between the purified signal and the contaminated observation. Typically, the output of a filter is a linear function of current and perhaps past observations. Some control charts employ linear filters. A Shewhart individual chart uses the original process data, while

an Exponentially Weighted Moving Average (EWMA) chart (Roberts (1959)) filters process data using a weighted average of past and current observations. Statistics of other popular control charts, such as the Cumulative Sum (CUSUM) chart (Page (1954)), the combined Shewhart-CUSUM chart (Lucas (1982)), the optimal EWMA (Srivastava and Wu (1993, 1997), Wu (1994)), the generalized likelihood ratio (Siegmund (1985), Siegmund and Venkatraman (1995), Apley and Shi (1994, 1999)), the adaptive CUSUM (Sparks (2000)), the adaptive EWMA (Capizzi and Masarotto (2003)), and the generalized EWMA (Han and Tsung (2004)) invoke linear combinations of the data process. The ARMA chart (Jiang, Tsui and Woodall (2000)) and the PID chart (Jiang et al. (2002)) use more complex forms to combine past and current information for filtering.

Optimal design of a linear filter for statistical process control was proposed by Chin and Apley (2007). The underlying ARMA process is identified by applying an impulse signal; coefficient estimation is achieved and the optimal design is created based on the assumed noise type, the charting method, and the desired in-control average run length (*ARL*). Simulation shows that this control chart can outperform existing control charts in some situations.

Although the linear filter-type control chart provides a good way to handle certain kinds of noise, it has limitations. For instance, the optimal design of a linear filter given by Chin and Apley (2007) depends heavily on the assumed shift magnitude. The magnitude of the shift is, however, rarely pre-specified, but is rather an unknown shift in a possible range, or even a dynamic shift that changes over time. Thus, a linear filter may achieve the optimum performance at a single point, but not over the mean shift range of interest. Since linear filters are designed for detecting certain types of noise, if the actual noise encountered is different or more complex, the performance of the filter can easily suffer.

To conquer such limitations, we propose extending conventional control charts that use linear filters to charts that apply nonlinear filters to the data. We expect that in presence of more complicated noise, the proposed nonlinear filter control charts can prove more capable of handling unknown and dynamic mean shifts, that is, it may respond to mean shifts more quickly. The most popular criterion for evaluating the responsiveness of a control chart is average run length (*ARL*): the average number of samples (subgroups) taken before an alarming signal is given.

In the next section, a nonlinear filter control chart is presented. A theorem on an estimation of the average run length, a discussion on the theoretical comparison of any two control charts, and the optimal design of an NFC chart, are given in Section 3. Section 4 gives a comparison of detection performance of the nonlinear filter charts and traditional control schemes, by numerical simulation. The definition and settings of two adaptive CUSUM charts

are shown in Appendix I. The proof of Theorem 1 in Section 3 can be found at <http://www.stat.sinica.edu.tw/statistica> as an on-line supplement.

2. A Nonlinear-Filter Control Chart

Here process data are first filtered by a nonlinear filter before being drawn on a chart. We design a nonlinear filter control chart by recalling the well-known CUSUM chart before extending it.

Let X_i ($i = 1, 2, \dots$) be the i th observation from an *i.i.d.* process. Suppose at time τ , called a change point, the mean of X_i abruptly changes from μ_0 to μ ; Thus, from time τ and on, the mean of X_i undergoes a step mean shift, $\delta = \mu - \mu_0$. We assume μ and τ are unknown, and that μ_0 and the standard deviation σ of the process $\{X_i, i \geq 1\}$ are known. Without loss of generality, $\mu_0 = 0$ and $\sigma = 1$.

The first time (stopping time) outside the control limit, $c > 0$, for the one-sided CUSUM chart, T , can be written as

$$T(\delta) = \inf \left\{ n : \max_{1 \leq k \leq n} \left[\sum_{i=n-k+1}^n \delta \left(X_i - \frac{\delta}{2} \right) \right] \geq c \right\}, \tag{2.1}$$

where $\delta/2 > 0$ is the reference value related to the magnitude of the mean shift, δ , see Hawkins and Olwell (1998).

Moustakides (1986) and Ritov (1990) have shown that the performance of the one-sided CUSUM control chart with a reference value of $\delta/2$ is optimal if the real mean shift is δ . In fact, however, we rarely know the exact magnitude of future mean shifts. To detect an unknown mean shift quickly, we may consider replacing δ with $\{|X_i|\}$ in the CUSUM chart, since $\{|X_i|\}$ contains real-time information about the magnitude of the mean shift. Thus, we take the stopping time of a new control chart to be:

$$T = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} \left[\sum_{i=n-k+1}^n |X_i| \left(X_i - \frac{|X_i|}{2} \right) \right] \geq c \right\}. \tag{2.2}$$

Here, the control statistic, $\sum_{i=n-k+1}^n |X_i|(X_i - |X_i|/2)$, is a nonlinear combination of the observed data, $\{X_i\}$. Since

$$|x| \left(x - \frac{|x|}{2} \right) = \begin{cases} \frac{x^2}{2} & \text{if } x \geq 0 \\ -\frac{3x^2}{2} & \text{if } x < 0, \end{cases}$$

we obtain the stopping time of a more general chart,

$$T(f_\alpha) = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} \left[\sum_{i=n-k+1}^n f_\alpha(X_i) \right] \geq c \right\}, \tag{2.3}$$

by taking a nonlinear function,

$$f_\alpha(x) = \begin{cases} \frac{x^\alpha}{2} & \text{if } x \geq 0 \\ -\frac{3|x|^\alpha}{2} & \text{if } x < 0, \end{cases} \quad (2.4)$$

where $\alpha > 0$. We call $T(f_\alpha)$ and $f_\alpha(\cdot)$, respectively, the stopping times of a nonlinear filter control chart (NFC) and a nonlinear filter function.

Generally, the nonlinear filter is not restricted to the form of (2.4). It can take any form. When f is a linear function, we call $T(f)$ the stopping time of a linear filter control chart (LFC).

If a negative mean shift is of interest, the stopping time of a general NFC chart can be written as

$$T(f_\alpha) = \inf \left\{ n \geq 1 : \min_{1 \leq k \leq n} \left[\sum_{i=n-k+1}^n f_\alpha(X_i) \right] \leq -c \right\}, \quad (2.5)$$

where $f_\alpha(x)$ is

$$f_\alpha(x) = \begin{cases} \frac{3x^\alpha}{2} & \text{if } x \geq 0 \\ -\frac{|x|^\alpha}{2} & \text{if } x < 0. \end{cases} \quad (2.6)$$

Thus, a two-sided NFC control scheme can be readily constructed.

Han and Tsung (2006) proposed a reference-free Cuscore (RFCuscore) chart that is also capable of tracing and detecting dynamic mean changes quickly without knowing the reference pattern or having prior knowledge of the mean shift magnitude. The RFCuscore chart is a special case of the nonlinear filter control chart with $\alpha = 2$.

3. Theoretical Analysis

For convenience of discussion, we use the standard quality control terminology. Let $P(\cdot)$ and $E(\cdot)$ denote the probability and expectation functions when there is no change in the mean, $\mu = \mu_0 = 0$; let $P_\mu(\cdot)$ and $E_\mu(\cdot)$ be the probability function and expectation function when the change point is at $\tau = 1$, and the true mean shift value is $\mu \neq 0$. The two most frequently used operating characteristics of statistical control charts are in-control average run length $ARL_0(T) = E(T)$ and out-of-control average run length $ARL_\mu(T) = E_\mu(T)$.

For comparison, all candidate charts share the same ARL_0 which corresponds to the same level of type I error rate. The chart with the smallest out-of-control ARL_μ at the desired mean shift magnitude has the highest power to detect the pre-specified shift.

In this section, we first present a theorem on the estimation of the *ARL* for a general control chart, including NFC and LFC, with a large control limit. Then we discuss how to compare the detection performances of any two control charts.

3.1. Approximation, estimation, and comparison of *ARL*'s

Let Y_1, Y_2, \dots be *i.i.d.* observations of a random variable Y , and let F and E be its cumulative distribution function and expectation function. Suppose F satisfies the following.

(I) The moment-generating function $h(\theta) = E(e^{\theta Y}) < \infty$ for some $\theta > 0$.

(II) For $x > E(Y)$ there is a $\theta(x) \in (0, \theta_1)$ such that $x = h'(\theta(x))/h(\theta(x))$, where $\theta_1 = \sup\{\theta : h(\theta) < \infty\}$.

Let $E(Y) < 0$. Since $h'(0) = E(Y) < 0$, $h'(\theta)/h(\theta)$ is strictly increasing (see Durrett (1991, p.60)) and $\log h(\theta) \rightarrow +\infty$ as $\theta \rightarrow \theta_1$, it follows that there exists at most one $\theta^* \in (\theta(0), \theta_1)$ such that $h(\theta^*) = 1$ or $\log h(\theta^*) = 0$, where $\theta(0) > 0$ satisfies $0 = h'(\theta(0))/h(\theta(0))$. That is, $h(\theta)$ attains its minimum value at $\theta(0) > 0$. We can call θ^* an **exponential rate** of Y . The meaning of θ^* is given in Theorem 1.

Now we define the stopping time of a control chart as

$$T = \inf \left\{ n : \max_{1 \leq k \leq n} \left[\sum_{i=n-k+1}^n Y_i \right] \geq c \right\}, \tag{3.1}$$

where $c > 0$ is the control limit.

Theorem 1. *Suppose the conditions (I) and (II) hold. If $E(Y_i) < 0$, then*

$$E(T) \sim D(c)e^{c\theta^*} \tag{3.2}$$

for large c , where $\theta^* > 0$ is the exponential rate satisfying $h(\theta^*) = 1$, $1/bc \leq D(c) \leq c/u$, $u = h'(\theta^*) > 0$ and b is a positive constant. If $E(Y_i) > 0$, then for large c ,

$$E(T) \sim \frac{c}{E(Y_i)}. \tag{3.3}$$

Here we do not consider the case that $E(Y_i) = 0$ since it is then difficult to estimate the *ARL*. The proof of Theorem 1 is in Section 3. The on-line supplement is at www.stat.sinica.edu.tw. Table 3.1 shows the out-of-control *ARL*s of NFC charts having $\alpha = 1.0$ and $\alpha = 2.0$ in the presence of mean shifts of magnitude around θ^* . The results were obtained via Monte Carlo simulation. The values of θ^* of the two control charts are highlighted in boldface. We can see that, although the formulas for $E(T)$ have different forms around θ^* , the out-of-control *ARL* evolves quite smoothly there.

Table 3.1. Out-of-control ARLs of two NFC charts with $\alpha = 1.0$ and $\alpha = 2.0$ around θ^* , in-control ARL =700.

$\alpha = 1.0, c = 5.148$		$\alpha = 2.0, c = 10.295$	
δ	ARL	δ	ARL
0.40000	59.8060	0.30000	94.8590
0.43600	50.2066	0.34300	74.4490
0.43610	50.1686	0.34310	74.1673
0.43620	50.1354	0.34320	74.1556
0.43625	50.0580	0.34330	74.0524
0.43630	50.0572	0.34340	74.0519
0.43635	50.0454	0.34350	73.9762
0.43640	50.0249	0.34355	73.9576
0.43650	50.0119	0.34360	73.9544
0.43660	49.9835	0.34365	73.9363
0.43670	49.9804	0.34370	73.9277
0.43680	49.9057	0.34380	73.7840
0.43690	49.8995	0.34390	73.6612
0.45000	47.0550	0.35000	71.2200
0.50000	37.9390	0.40000	56.3860

Remark 1. Large c means that the term $o(1)$ in $E(T) = (c/E(Y_i))(1 + o(1))$ of Theorem 1 is negligible. For example, if $c \geq 10$, then $|o(1)| < 1/c \leq 1/10 = 0.1$ for the CUSUM chart with $\delta = 1$ and $Y_i \sim N(1, 1)$. As can be seen, the results of Theorem 1 can be used for many control charts in detecting the observed process that is not necessarily normal. If $\{Y_i\}$ is normal and $E(Y) < 0$, then the exponential rate is $2|E(Y_i)|/\text{Var}(Y_i)$, which can be considered as signal-to-noise ratio. In fact, we write

$$h(\theta) = E(e^{\theta Y}) = \exp \left\{ \frac{\theta \text{Var}(Y)[\theta \text{Var}(Y) + 2E(Y_i)]}{2\text{Var}(Y)} \right\}. \tag{3.4}$$

Thus $h(\theta^*) = 1$ when $\theta^* = -2E(Y_i)/\text{Var}(Y_i) = 2|E(Y_i)|/\text{Var}(Y_i)$.

Theorem 1 can be considered as a generalization of Basseville and Nikiforov (1993, p.162), since $Y = \log[p_\theta(X)/p_{\theta_0}(X)]$ where $p_\theta(\cdot)$ and $p_{\theta_0}(\cdot)$ are two distribution functions or two density functions. As an application of Theorem 1, we discuss how to compare the detection performance of two control charts.

Let $X_i, i = 1, 2, \dots$ be an *i.i.d.* process with mean $E_\mu(X_i) = \mu \geq 0$. We say that there is no mean shift if $\mu = 0$. For any two linear or nonlinear filter functions, $f_j(\cdot), j = 1, 2$, we can get two control charts with stopping times $T(f_j), j = 1, 2$. Let $F_j(x)$ be the distribution function of $f_j(X_i), j = 1, 2$, and let $E_j(\mu) = E_\mu(f_j(X_i)), j = 1, 2$. Assume that conditions (I) and (II) hold for $F_j(x), j = 1, 2$, and that $E_j(\mu), j = 1, 2$ are strictly increasing functions on $\mu \geq 0$

with $E_j(0) < 0$. Thus, there exist two positive numbers, μ_1^* and μ_2^* , such that $E_j(\mu_j^*) = 0, j = 1, 2$. For a given $\mu < \mu_j^*$, let $\theta_j^*(\mu)$ be the exponential rate of $F_j(x)$ for $j = 1, 2$.

Suppose $E_0(T(f_1)) = E_0(T(f_2)) = ARL_0$. It follows from (3.2) of Theorem 1 that $D_1(c_1)e^{c_1\theta_1^*(0)} \sim D_2(c_2)e^{c_2\theta_2^*(0)}$ for large c_1 and c_2 , where c_1 and c_2 are control limits, respectively, for $T(f_1)$ and $T(f_2)$. Thus, $c_1\theta_1^*(0) = c_2\theta_2^*(0)(1 + o(1))$ for large c_1 and c_2 .

Let the mean shift be $\mu < \min\{\mu_1^*, \mu_2^*\}$. By (3.2) of Theorem 1, we have

$$\begin{aligned} ARL_\mu(T(f_1)) &= E_\mu(T(f_1)) = (1 + o(1))D_1(c_1) \exp\{c_1\theta_1^*(\mu)\} \\ &= (1 + o(1))D_1(c_1) \exp\left\{c_2 \frac{\theta_2^*(0)\theta_1^*(\mu)}{\theta_1^*(0)}(1 + o(1))\right\}, \\ ARL_\mu(T(f_2)) &= E_\mu(T(f_2))(1 + o(1))D_2(c_2) \exp\{c_2\theta_2^*(\mu)\}. \end{aligned}$$

From this, we see that $ARL_\mu(T(f_1)) > ARL_\mu(T(f_2))$ for large c_1 and c_2 if and only if

$$\frac{\theta_1^*(\mu)}{\theta_1^*(0)} > \frac{\theta_2^*(\mu)}{\theta_2^*(0)}. \tag{3.5}$$

Similarly, for $\mu > \max\{\mu_1^*, \mu_2^*\}$, we get from (3.3) of Theorem 1 that $ARL_\mu(T(f_1)) > ARL_\mu(T(f_2))$ for large c_1 and c_2 if and only if

$$\frac{1}{\theta_1^*(0)E_1(\mu)} > \frac{1}{\theta_2^*(0)E_2(\mu)}. \tag{3.6}$$

Suppose $\mu_1^* < \mu_2^*$. Then $ARL_\mu(T(f_1)) < ARL_\mu(T(f_2))$ for $\mu_1^* < \mu < \mu_2^*$ and large c_1 and c_2 . In fact, by Theorem 1, we have

$$ARL_\mu(T(f_1)) = (1 + o(1))\frac{c_1}{E_1(\mu)} = (1 + o(1))\frac{c_2\theta_2^*(0)}{\theta_1^*(0)E_1(\mu)},$$

since $E_1(\mu) = E_\mu(f_1(X_i)) > 0$ for $\mu > \mu_1^*$, and

$$ARL_\mu(T(f_2)) = (1 + o(1))D_2(c_2) \exp\{c_2\theta_2^*(\mu)\}$$

for $\mu < \mu_2^*$. Thus, $ARL_\mu(T(f_1)) < ARL_\mu(T(f_2))$ for large c_2 . Similarly, if $\mu_1^* > \mu_2^*$, $ARL_\mu(T(f_1)) > ARL_\mu(T(f_2))$ holds for $\mu_1^* > \mu > \mu_2^*$ and large c_1 and c_2 .

These comparisons are applied to the specific NFC examples that are derived from conventional CUSUM charts in the next subsection.

3.2. Examples

Let $\{X_i, i \geq 1\}$ be *i.i.d.* $N(\mu, 1)$, $\mu \geq 0$. Denote the expectation and variance of $g_\delta(X_1)$, $f_1(X_1)$ and $f_2(X_1)$, respectively, by $E_0(\mu)$, $V_0(\mu)$, $E_1(\mu)$, $V_1(\mu)$, and $E_2(\mu)$, $V_2(\mu)$, where $g_\delta(x) = \delta(x - \delta/2)$, and $f_i(\cdot)$, $i = 1, 2$, are the functions $f_\alpha(\cdot)$ with $\alpha = 1, 2$ defined in (2.4). Then

$$\begin{aligned} E_0(\mu) &= E(g_\delta(X_1)) = \delta(\mu - \frac{\delta}{2}), & V_0(\mu) &= \text{Var}(g_\delta(X_1)) = \delta^2, \\ E_1(\mu) &= \mu[\frac{3}{2} - \Phi(\mu)] - \varphi(\mu), \\ V_1(\mu) &= \frac{9}{4} + \mu^2\Phi(\mu)[1 - \Phi(\mu)] + \varphi(\mu)(\mu - \varphi(\mu)) - 2\Phi(\mu)(1 + \mu\varphi(\mu)), \\ E_2(\mu) &= 2\mu\varphi(\mu) + (\mu^2 + 1)[2\Phi(\mu) - \frac{3}{2}], \\ V_2(\mu) &= 4\Phi(\mu)(\mu^2 + 1)^2[1 - \Phi(\mu)] + \mu^2(4\mu\varphi(\mu) + 9) + 4\Phi(\mu) + \frac{9}{2} \\ &\quad - 4(\mu\varphi(\mu) + 1)(\mu\varphi(\mu) + 2\Phi(\mu)(\mu^2 + 1)), \end{aligned}$$

where $\Phi(\cdot)$ and $\varphi(\cdot)$ are the standard normal distribution and density functions. It can be shown that $E_j(\mu)$, $j = 0, 1, 2$, are all strictly monotonically increasing functions on $\mu \geq 0$, with $E_0(0) = -\delta/2$, $E_0(\delta/2) = 0$, $E_1(0) = -1/\sqrt{2\pi}$, $E_1(0.4363) = 0$, and $E_2(0) = -1/2$, $E_2(0.3436) = 0$. Denote the moment-generating functions and the exponential rates of $g_\delta(X_i)$ and $f_j(X_i)$, $j = 1, 2$, respectively, by $h_0(\theta)$ and $\theta_0^*(\mu)$, and $h_j(\theta)$ and $\theta_j^*(\mu)$, $j = 1, 2$. We then get

$$\begin{aligned} h_0(\theta) &= E(e^{\theta g_\delta(X_1)}) = \exp\left\{\frac{\theta\delta^2(\theta\delta^2 + 2\delta(\mu - \delta/2))}{2\delta^2}\right\}, \\ h_1(\theta) &= E(e^{\theta f_1(X_1)}) = \exp\left\{\frac{4\mu\theta + \theta^2}{8}\right\}\Phi\left(\mu + \frac{\theta}{2}\right), \\ &\quad + \exp\left\{\frac{12\mu\theta + 9\theta^2}{8}\right\}\left[1 - \Phi\left(\mu + 3\frac{\theta}{2}\right)\right], \\ h_2(\theta) &= E(e^{\theta f_2(X_1)}) = \frac{1}{\sqrt{1-\theta}} \exp\left\{\frac{\theta\mu^2}{2(1-\theta)}\right\}\Phi\left(\frac{\mu}{\sqrt{1-\theta}}\right) \\ &\quad + \frac{1}{\sqrt{3\theta+1}} \exp\left\{\frac{-3\theta\mu^2}{2(3\theta+1)}\right\}\left[1 - \Phi\left(\frac{\mu}{\sqrt{3\theta+1}}\right)\right]. \end{aligned}$$

Obviously, $h(\theta_0^*(\mu)) = 1$ for $\theta_0^*(\mu) = (\delta - 2\mu)/\delta = |2E(g_\delta(X_i))|/\text{Var}(g_\delta(X_i))$ when $0 \leq \mu \leq \delta/2$. It is rare for the exponential rate $\theta_j^*(\mu)$, $j = 1, 2$, to have a closed form, but $\theta_j^*(\mu)$, $j = 1, 2$, strictly monotonically decrease on $0 \leq \mu \leq 0.4363$ and $0 \leq \mu \leq 0.3436$, respectively, with $\theta_1^*(0) = 1.04$, $\theta_1^*(0.4363) = 0$, $\theta_2^*(0) = 0.449$ and $\theta_2^*(0.3436) = 0$. Note that $E_1(0) = -1/\sqrt{2\pi}$, $V_1(0) = 5/4 - 1/2\pi$,

$E_2(0) = -1/2, V_2(0) = 7/2$, and therefore $\theta_1^*(0) = 1.04 \neq 2|E_1(0)|/V_1(0)$ and $\theta_2^*(0) = 0.449 \neq 2|E_2(0)|/V_2(0)$.

Using Theorem 1, we can get the approximate ARL 's for the CUSUM chart with stopping time $T(g_\delta)$, and the NFC charts with stopping times, $T(f_j), j = 1, 2$. Thus

$$ARL_\mu(T(g_\delta)) \sim D_0(c)e^{c(\delta-2\mu)/\delta}, \quad ARL_\mu(T(g_\delta)) \sim \frac{c}{\delta(\mu - \delta/2)},$$

respectively, for $0 \leq \mu < \delta/2$ and $\mu > \delta/2$, and

$$ARL_\mu(T(f_j)) \sim D_j(c)e^{c\theta_j^*(\mu)}, \quad ARL_\mu(T(f_j)) \sim \frac{c}{E_j(\mu)},$$

respectively, for $0 \leq \mu < \mu_j^*$ and $\mu > \mu_j^*, j = 1, 2$, where $\mu_1^* = 0.4363$ and $\mu_2^* = 0.3436$.

Note that the ARL s for the CUSUM chart, $ARL_\mu(T(g_\delta))$, are consistent with the known results when the control limit is large (see Srivastava and Wu (1997)).

Suppose that the control limits of $T(f_1), T(f_2)$, and $T(g_\delta)$ have a large common in-control ARL_0 . By checking (3.5) and (3.6) for the three charts, we can make comparisons.

Remark 2. (1) $ARL_\mu(T(f_1)) < ARL_\mu(T(f_2))$ for $0.6597 < \mu < 1.7162$, and $ARL_\mu(T(f_1)) > ARL_\mu(T(f_2))$ for $0 < \mu < 0.6597$ and $\mu > 1.7162$.

(2) For $\delta = 1, ARL_\mu(T(g_\delta)) > ARL_\mu(T(f_1))$ for $0 < \mu < 0.7645, ARL_\mu(T(g_\delta)) < ARL_\mu(T(f_1))$ for $\mu > 0.7645, ARL_\mu(T(g_\delta)) > ARL_\mu(T(f_2))$ for $0 < \mu < 0.7164$ and $\mu > 3.5437$, and $ARL_\mu(T(g_\delta)) < ARL_\mu(T(f_2))$ for $0.7164 < \mu < 3.5437$.

The one-sided CUSUM control chart with a reference value of $\delta/2$ is optimal: $ARL_\mu(T(g_\delta))$ is the smallest if the real mean shift is δ . It follows from Remark 2 (2) that both $ARL_\mu(T(f_1))$ and $ARL_\mu(T(f_2))$ are shorter than $ARL_\mu(T(g_\delta))$ when the real mean shift μ satisfies $0 < \mu < 0.7645 < \delta = 1$ and $\mu > 3.5437 > \delta = 1$, respectively. The optimality of the CUSUM control chart thus depends on the choice of reference value $\delta/2$.

Although these theoretical comparisons are based on the condition that the control limits, and therefore the common ARL_0 , are large, the simulation results in the next section produce results consistent with Remark 2.

Remark 3. Take the nonlinear filter f_α defined in (2.4) and let $\{Y_i(\alpha) = f_\alpha(X_i), i \geq 1\}$, where $\{X_i, i \geq 1\}$ is *i.i.d.* normal. We have $E(e^{\theta Y_1(\alpha)}) = +\infty$ for $\alpha > 2$ and any $\theta > 0$. That is, condition (I) and therefore result (3.2) in Theorem 1 do not hold for $\{Y_i(\alpha), i \geq 1\}$ when $\alpha > 2$. Thus, (3.5) and (3.6) cannot be

used for the case of $\alpha > 2$. However, the results in the next section show that the two nonlinear filter control charts with stopping times $T(f_\alpha)$ for $\alpha = 2.5$ and $\alpha = 3$, respectively, have good detection performance in the presence of dynamic mean shifts.

3.3. Design of an optimal NFC chart

The nonlinear filter control chart provides great flexibility in functional forms, although the optimal design of such a control chart is challenging. Theoretically, any function could be a candidate for the optimal choice. Here, we only consider filter functions in

$$D_0 = \{f : E_\mu(f(X_i)) \text{ satisfies conditions (I) and (II) with } E_0(f(X_i)) < 0\}.$$

If a NFC chart, defined at (2.3) with $f^*(\cdot)$ in D_0 , has the best performance in detecting the unknown mean change in the range $(0, R)$, where $0 < R \leq \infty$, we call it the optimal filter function. Based on Section 3.1 and the inequalities (3.5) and (3.6), we can conclude that the exponential rate and the expectation of the optimal filter function, $f^*(\cdot)$ satisfy

$$\begin{aligned} & a \int_0^{\mu_f \wedge R} \frac{\theta_{f^*}^*(\mu)}{\theta_{f^*}^*(0)} d\mu + b \int_{(\mu_f + \epsilon) \wedge R}^R \frac{1}{\theta_{f^*}^*(0) E_\mu(f^*(X_1))} d\mu \\ &= \min_{f \in D_0} \left\{ a \int_0^{\mu_f \wedge R} \frac{\theta_f^*(\mu)}{\theta_f^*(0)} d\mu + b \int_{(\mu_f + \epsilon) \wedge R}^R \frac{1}{\theta_f^*(0) E_\mu(f(X_1))} d\mu \right\}, \end{aligned} \tag{3.7}$$

where a, b , are two given positive constants with $a + b = 1$, ϵ is a small positive number, μ_f satisfies $E_{\mu_f}(f(X_1)) = 0$, $E_\mu(f(X_1)) < 0$ for $0 \leq \mu < \mu_f$, and $E_\mu(f(X_1)) > 0$ for $\mu > \mu_f$. Here, a and b represent, respectively, the weight of “small mean change, $0 < \mu < \mu_f$ ” and “large mean change, $\mu > \mu_f$ ”. The optimal filter function $f^*(\cdot)$ defined in (3.7) usually depends on a, b, ϵ and R .

While (3.7) provides a unified criterion for selecting an optimal control chart for either NFC or LFC, it is not easy to obtain the optimal NFC chart since the functions $\theta_f^*(\mu)/\theta_f^*(0)$ and $\theta_f^*(0)E_\mu(f(X_i))$ usually have no closed form. If the set D_0 could be reduced, obtaining the optimal NFC chart becomes possible. For example, let $\{X_i, i \geq 1\}$ be *i.i.d.* $N(\mu, 1)$, where $\mu \geq 0$. Let

$$D_0 = D_0(R, \epsilon) = \{f_\delta : f_\delta(x) = \delta(x - \delta/2), |x| < \infty, 0 < \delta < 2(R - \epsilon)\},$$

where $R > 0, \epsilon > 0$. Obviously, for any fixed $\delta > 0$, $f_\delta(X_1)$ satisfies conditions (I) and (II). Since $\theta_{f_\delta}^*(\mu) = (\delta - 2\mu)/\delta$ for $0 \leq \mu < \delta/2$, $\mu_{f_\delta} = \delta/2$ and $E_\mu(f_\delta(X_1)) = \delta(\mu - \delta/2)$ for $\mu > \delta/2$, it follows that

$$a \int_0^{\mu_f \wedge R} \frac{\theta_f^*(\mu)}{\theta_f^*(0)} d\mu + b \int_{(\mu_f + \epsilon) \wedge R}^R \frac{1}{\theta_f^*(0) E_\mu(f(X_1))} d\mu$$

$$\begin{aligned}
 &= a \int_0^{\delta/2} \left(1 - \frac{2\mu}{\delta}\right) d\mu + b \int_{(\delta/2+\epsilon)\wedge R}^R \frac{1}{\delta(\mu - \delta/2)} d\mu \\
 &= \frac{a\delta}{4} + \frac{b}{\delta} \left[\ln\left(R - \frac{\delta}{2}\right) - \ln \epsilon\right] \triangleq H(\delta).
 \end{aligned}$$

We further have $H'(0) = -\infty$, $H'(2(R - \epsilon)) = a/4 - b/(4\epsilon(R - \epsilon))$, and

$$H''(\delta) = \frac{2b}{\delta^3} \left[\ln\left(R - \frac{\delta}{2}\right) - \ln \epsilon\right] + \frac{b(4R - \delta)}{\delta^2(2R - \delta)^2} > 0$$

for $0 < \delta < 2(R - \epsilon)$. If $a\epsilon(R - \epsilon) \leq b$, then $H(2(R - \epsilon)) = \inf_{0 \leq \delta \leq 2(R - \epsilon)} H(\delta)$. Thus, the optimal filter function in $D_0(R, \epsilon)$ is $f^*(x) = (2R - \epsilon)(x - R - \epsilon)$. If $a\epsilon(R - \epsilon) > b$, then $H(\delta^*) = \inf_{0 \leq \delta \leq 2(R - \epsilon)} H(\delta)$, and therefore, the optimal filter function is $f^*(x) = \delta^*(x - \delta^*/2)$, where δ^* is the unique solution to

$$H'(\delta) = \frac{a}{4} - \frac{b}{\delta^2} \left[\ln\left(R - \frac{\delta}{2}\right) - \ln \epsilon\right] - \frac{b}{\delta(2R - \delta)} = 0$$

for $0 \leq \delta \leq 2(R - \epsilon)$.

4. Simulation and Comparison

In this section, we further demonstrate that the nonlinear filter control charts are superior at detecting dynamic mean shifts.

We chose nine control charts for comparison, a two-sided CUSUM chart with $\delta = 1$, six two-sided nonlinear filter charts with $\alpha = 1.0, 1.5, 1.8, 2.0, 2.5$, and 3.0 , denoted by NFC_α , and two adaptive CUSUM charts proposed by Sparks (2000) because they are specifically designed for detecting shifts within a range. According to the Sparks (2000)'s recommendation, we build two adaptive CUSUM charts, $ACUSUM1$ and $ACUSUM2$, for comparison purposes. The definitions and settings of the two adaptive CUSUM charts are shown in Appendix I.

Observations, $X_i, i \geq 1$, were *i.i.d.* normal $N(\mu, 1)$. The possible mean shift values were $\mu = 0.05, 0.1, \dots, 6$. As for the dynamic mean shifts, we assumed that $X_i, i \geq 1$, were mutually independent and $X_k \sim N(\mu p_k, 1)$, with four types of dynamic mean shifts $\{\mu p_k\}$ studied, respectively, in Figures 4.2–4.4. Figure 4.1 compares simulation results for step shifts in process mean, while Figures 4.2–4.5 contain results for four types of dynamic mean shifts. For dynamic mean shifts, the definition of the first time (stopping time) outside the control limit was the same as (2.3). Note that the results of Theorem 1 are not true for the stopping time of the dynamic mean shift, since $Y_k = f_\alpha(X_k), k \geq 1$, had different distributions, where $X_k \sim N(\mu p_k, 1)$ and f_α is defined in (2.4).

The numerical *ARLs* were obtained from 1,000,000 Monte Carlo simulations. Although *ARL* is a popular criterion, it is deficient in evaluating the charting

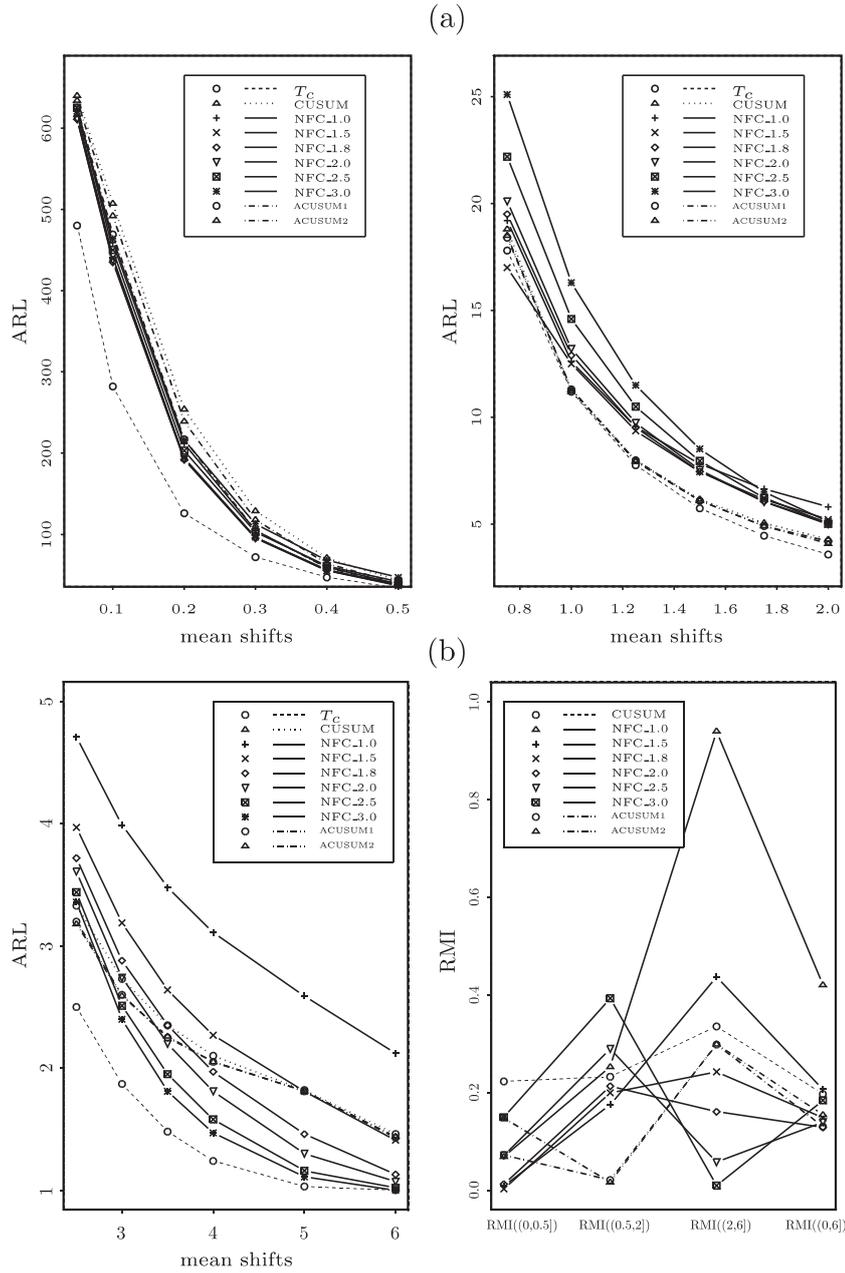


Figure 4.1. Comparison of ARL_μ 's of the nine control charts with $ARL_0 = 700$.

performance for a range of anticipated mean shifts. To handle such a situation, Han and Tsung (2006) proposed the relative mean index (RMI) for a control

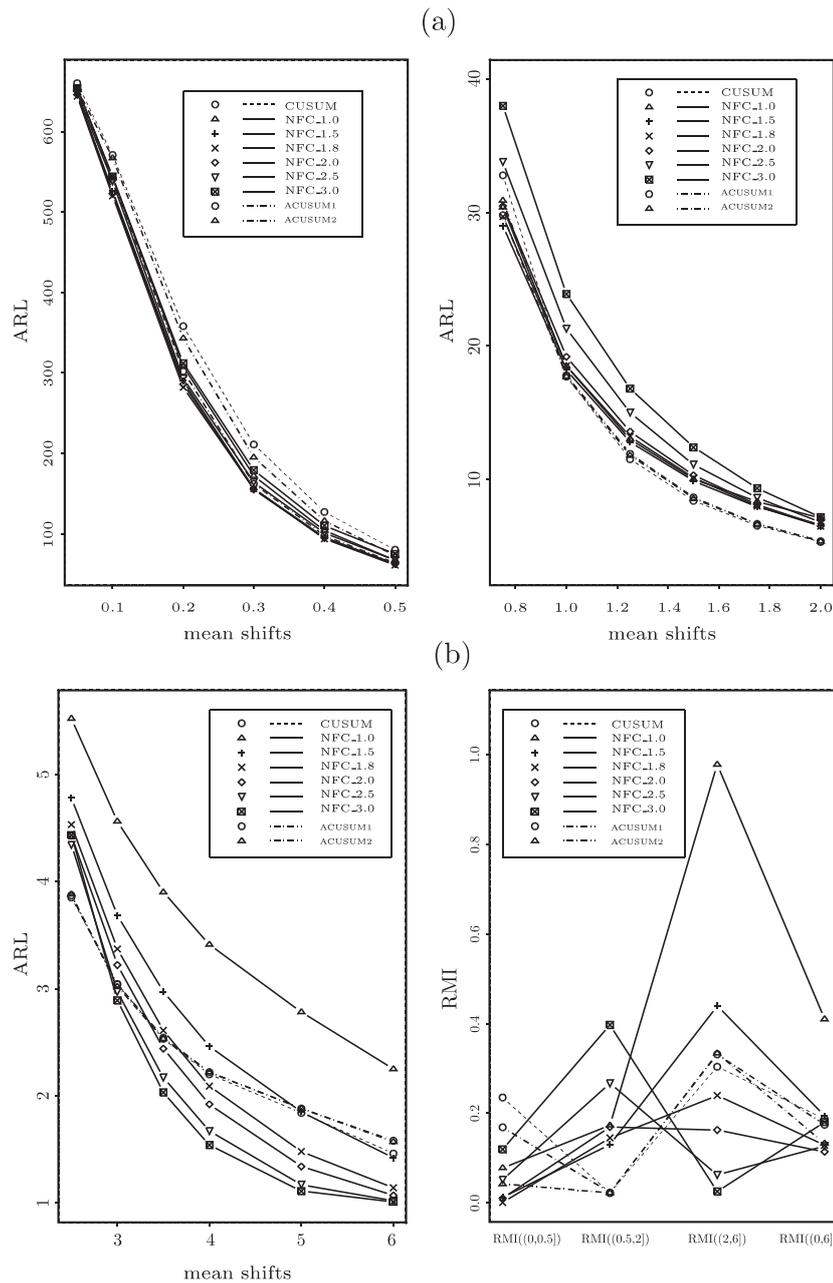


Figure 4.2. Comparison of $ARLs$ of the nine control charts with $ARL_0 = 700$, $p_k = 3/4 + 1/4(1/2)^{k-1}$.

chart as

$$RMI(a, b) = \frac{1}{n} \sum_{i=1}^n \left[\frac{ARL_{\mu_i}(T) - ARL_{\mu_i}^*}{ARL_{\mu_i}^*} \right], \quad (4.1)$$

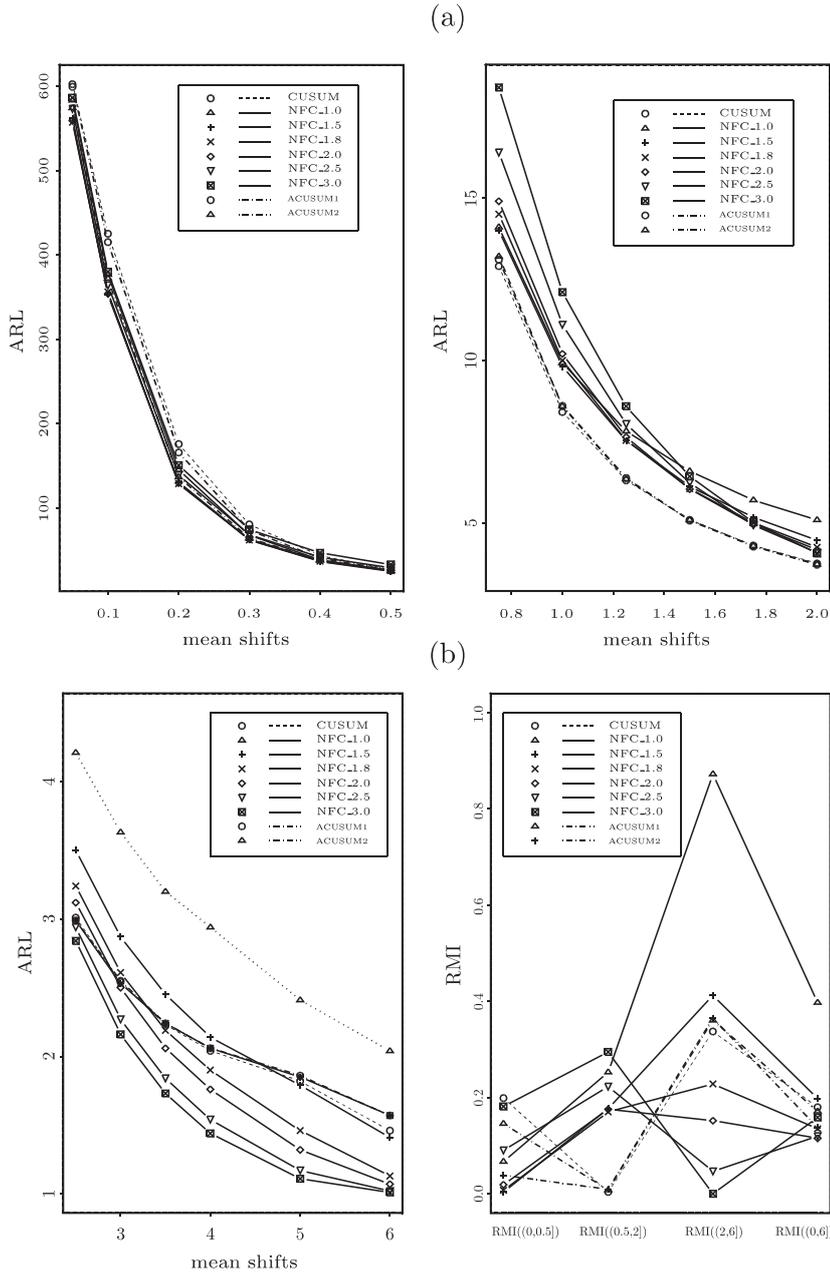


Figure 4.3. Comparison of ARL s of the nine control charts with $ARL_0 = 700$, $p_k = 5/4 - 1/4(1/2)^{k-1}$.

where μ_i and $1 \leq i \leq n$ are the mean shifts in the anticipated range $(a, b]$ ($0 < a < b$) within which the control chart performance is evaluated, $ARL_{\mu_i}(T)$

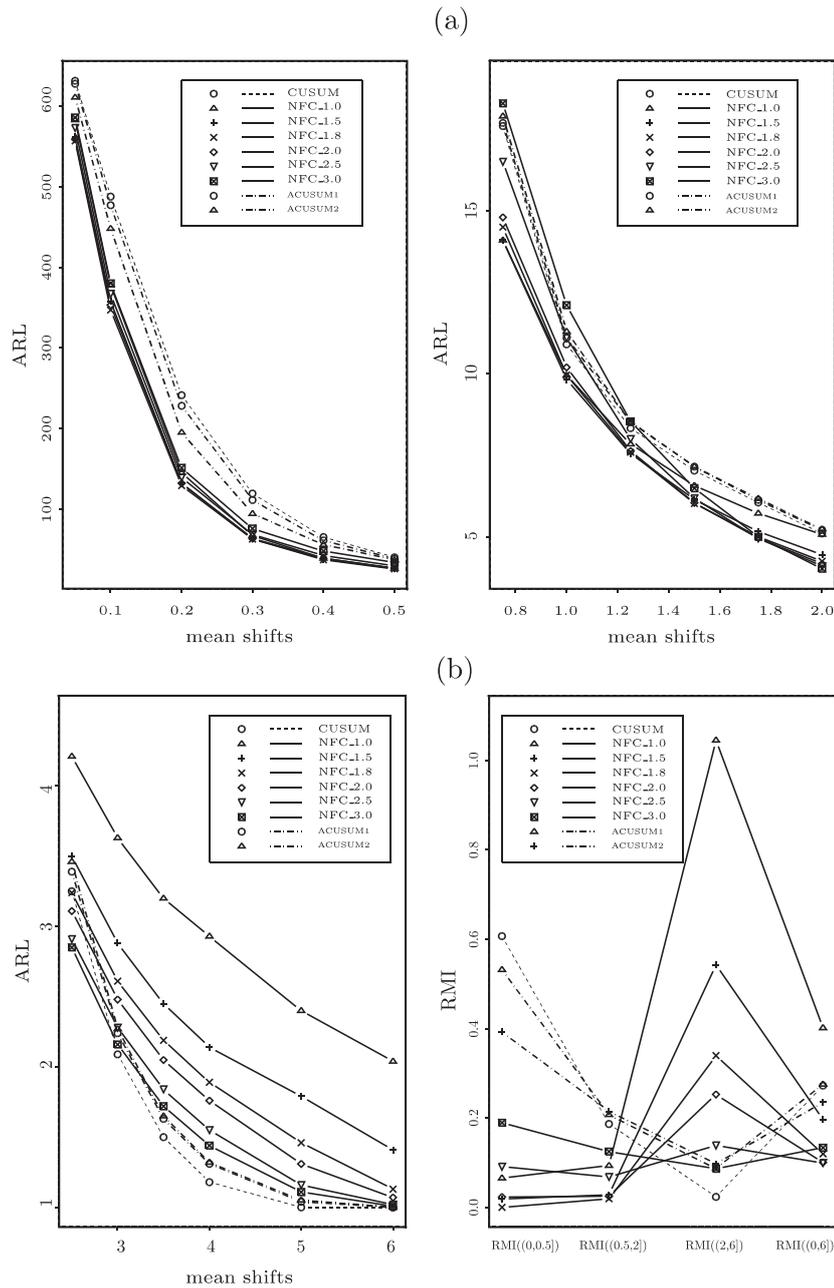


Figure 4.4. Comparison of ARL s of the nine control charts with $ARL_0 = 700$, $p_k = 1 + \cos(k\pi/4)$.

is the ARL of a control chart when the real mean shift is μ_i . Here $ARL^*_{\mu_i}$ denotes the smallest value of ARL of all control charts in comparison when the mean shift,

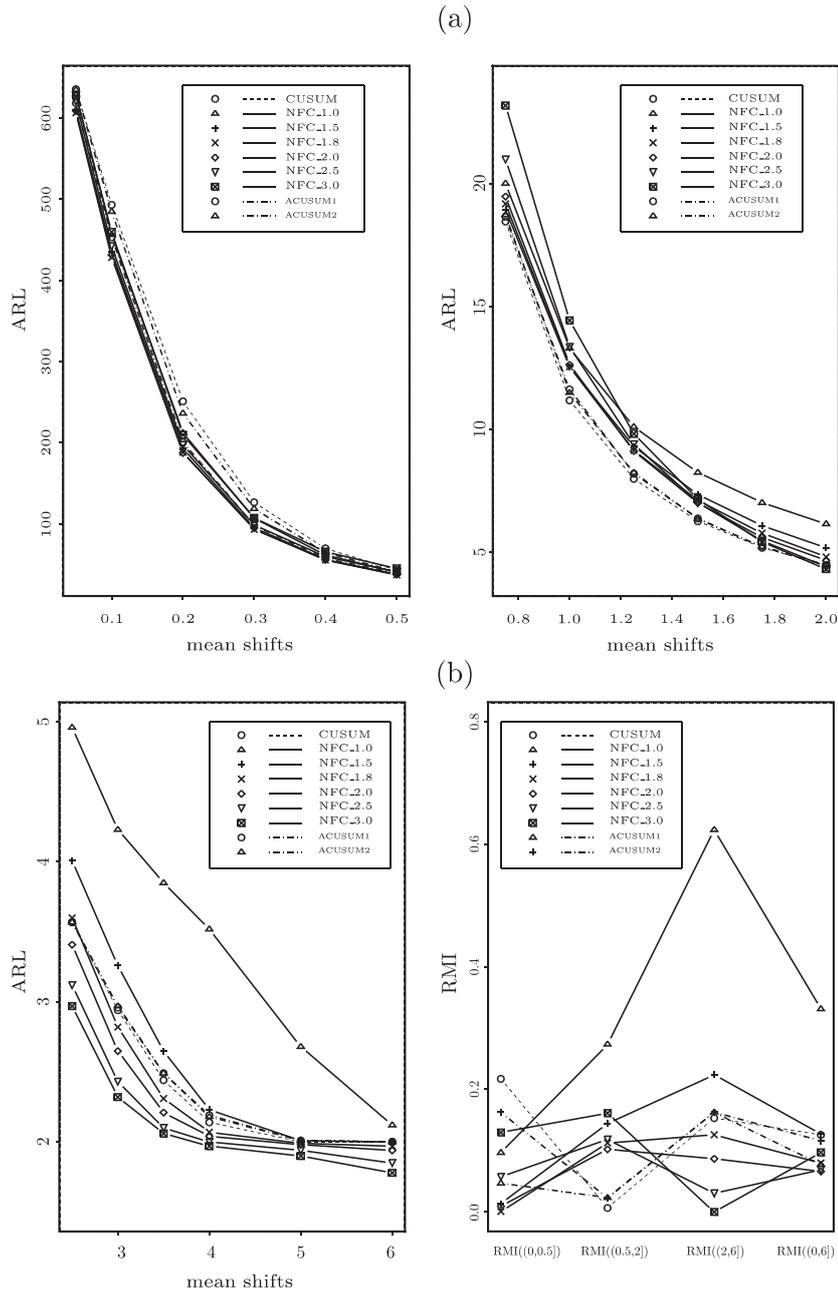


Figure 4.5. Comparison of $ARLs$ of the nine control charts with $ARL_0 = 700$, $p_k = 1 + (-1)^k/2$.

μ_i , occurs. Obviously, the smaller $RMI(T)$, the better the control chart is at

detecting mean shifts on the whole. Rather than using *ARL* at a specific shift magnitude as a criterion, the proposed *RMI* can take all the possible mean shifts within a range into consideration.

Figures 4.1–4.5 give the detailed numerical results of the out-of-control average run length. Here, all control charts were two-sided, and their common in-control ARL_0 was fixed at 700. $RMI((0, 0.5])$, $RMI((0.5, 2])$, $RMI((2, 6])$, and $RMI((0, 6])$ are the relative mean index (*RMI*) of the small, medium and large mean shifts, respectively. The smallest *RMI* values for a specific mean shift are highlighted in boldface.

We first compare the results in Remark 2 with the Monte Carlo simulation results of the NFC charts.

Remark 4. (1). In Figure 4.1, for step shifts in the process mean, we have $ARL_\mu(T(f_1)) < ARL_\mu(T(f_2))$ for $0.55 < \mu < 1.36$, and $ARL_\mu(T(f_1)) > ARL_\mu(T(f_2))$ for $0 < \mu < 0.55$ and $\mu > 1.36$.

(2). In the presence of step changes in the process mean, for $\delta = 1$, $\alpha = 1$, and $\alpha = 2$, we have $ARL_\mu(T_C(1)) > ARL_\mu(T(f_1))$ for $0 < \mu < 0.70$, $ARL_\mu(T_C(1)) < ARL_\mu(T(f_1))$ for $\mu > 0.70$, $ARL_\mu(T_C(1)) > ARL_\mu(T(f_2))$ for $0 < \mu < 0.64$ and $\mu > 3.02$, and $ARL_\mu(T_C(1)) < ARL_\mu(T(f_2))$ for $0.64 < \mu < 3.02$.

Compared with Remark 2, we conclude that the theoretical results for large control limits are basically consistent with the simulation results for $ARL_0 = 700$.

For step mean shifts, Bagshaw and Johnson (1975) showed that the optimal conventional CUSUM chart for detecting a mean shift μ is the CUSUM chart with $\delta = \mu/2$. Therefore, in Figure 4.1, besides the *ARLs* of the nine control charts, the *ARLs* of the conventional two-sided CUSUM control chart with stopping time T_C , with a reference value of $\mu_i/2$ when the real mean shift is μ_i , are also shown. Obviously, the NFC charts with $\alpha = 1.8$ and $\alpha = 3$, respectively, perform the best in detecting small and large mean shifts. The adaptive CUSUM chart, *ACUSUM2*, has higher capability in detecting medium mean shifts. However, with regard to $RMI((0, 6])$, the NFC chart with $\alpha = 2.0$ has better overall performance than the CUSUM chart and the adaptive CUSUM charts.

In Figures 4.2–4.5, we show the performance of six NFC charts, one conventional CUSUM chart, and two adaptive CUSUM charts in the presence of four types of dynamic changes. Figures 4.2 and 4.3 show the effects of two types of damping mean shifts, $(3/4 + 1/4(1/2)^{k-1})\mu_i$ and $(5/4 - 1/4(1/2)^{k-1})\mu_i$. Obviously, these mean shifts increase and decrease, and finally stabilize at $\mu_i/4$ and $5\mu_i/4$. In Figure 4.4, a cyclic mean shift, $(1 + \cos(k\pi/4))\mu_i$ is considered. Figure 4.5 shows how the control charts can detect “zigzag” mean shifts.

In Figure 4.2, the NFC chart with $\alpha = 1.8$ and the NFC chart with $\alpha = 3.0$, respectively, had the best performances in detecting small and large mean shifts.

With regard to medium mean shifts, the conventional CUSUM chart beats all other charts with its smallest RMI values. However, judging from $RMI((0, 6])$, the NFC chart with $\alpha = 2.0$ enjoyed the best overall performance. Similar conclusions can be drawn from Figure 4.3. The NFC charts still outperformed the other two types of control charts in detecting small, large and overall mean shifts.

In Figure 4.4, the advantage of the NFC control charts still remains. For small and medium mean shifts, nearly all of the NFC charts performed better than the competing control charts. The NFC chart with $\alpha = 1.8$ had the highest capability in detecting small and medium mean shifts. As to large mean shifts, the conventional CUSUM chart had the best performance. However, with respect to the overall performance, the NFC chart with $\alpha = 2.5$ was the best. Therefore, in detecting cyclic mean shifts, NFC charts appear preferable to the conventional CUSUM chart and the adaptive CUSUM charts.

With regard to “zigzag” mean shifts, Figure 4.5 shows that NFC charts can detect small and large mean shifts. The NFC chart with $\alpha = 3.0$ was the best for detecting large mean shifts, while the NFC chart with $\alpha = 1.8$ performed the best in the presence of small mean shifts. Although the conventional control chart was most suitable for detecting medium mean shifts, the NFC chart with $\alpha = 2.0$ was the best for detecting a mean shift over the range $(0, 6]$.

In summary, we found NFC charts more capable of detecting constant or dynamic shifts in process mean. In particular, for step mean changes, the NFC chart with $\alpha = 2.0$ had better overall detection capability than either CUSUM or adaptive CUSUM charts. For damping mean shifts, we recommend using the NFC control chart with $\alpha = 2.0$. For cyclic mean shifts, $(1 + \cos(k\pi/4))\mu_i$, two NFC control charts with $\alpha = 1.8$ and $\alpha = 2.5$ were better choices.

Acknowledgement

We thank the Editors and the referees for their valuable comments and suggestions that have improved both the content and presentation of this work. This research was supported by RGC Competitive Earmarked Research Grants 620606 and 620707.

Appendix I.

The control statistics of the two-sided adaptive CUSUM chart proposed by Sparks (2000) are

$$\hat{\delta}_t^U = \max(\alpha x_{t-1} + (1 - \alpha)\hat{\delta}_{t-1}^U, \delta_{\min}^U),$$

$$Z_t^U = \max\left(0, Z_{t-1}^U + \frac{[x_t - \hat{\delta}_t^U/2]}{h(\hat{\delta}_t^U)}\right),$$

$$\hat{\delta}_t^L = \min(\alpha x_{t-1} + (1 - \alpha)\hat{\delta}_{t-1}^L, \delta_{\max}^L),$$

$$Z_t^L = \min\left(0, Z_{t-1}^L + \frac{[x_t - \hat{\delta}_t^L/2]}{h(-\hat{\delta}_t^L)}\right),$$

where x_t is the original observation of the process, $\hat{\delta}_t^L$ and $\hat{\delta}_t^U$ are the one-step-ahead forecast of the mean shift at time t , δ_{\max}^L and δ_{\min}^U , respectively, denote the smallest downside and upside mean shifts that are of interest, $h(\delta)$ is the control limit of the conventional CUSUM chart when its reference value is $k = \delta/2$, and the in-control ARL is a pre-specified value, here it is 700. Z_t^L and Z_t^U are the plotted statistics. The control chart signals whenever $Z_t^L < -h_z$ or $Z_t^U > h_z$. The value of h_z is chosen to achieve a specified in-control ARL of the adaptive CUSUM chart.

Sparks (2000) recommended using $\alpha = 0.1$, $\delta_{\max}^L = -0.5$, $\delta_{\min}^U = 0.5$, $\hat{\delta}_1^L = -1$, $\hat{\delta}_1^U = 1$ for detecting smaller shifts, and $\alpha = 0.1$, $\delta_{\max}^L = -0.75$, $\delta_{\min}^U = 0.75$, $\hat{\delta}_1^L = -1$, $\hat{\delta}_1^U = 1$ for detecting larger shifts. We denote them, respectively, as *ACUSUM1* and *ACUSUM2*.

References

- Apley, D. W. and Shi, J. (1994). A statistical process control method for autocorrelated data using a GLRT. In *Proceeding of the International Symposium on Manufacturing Science and Technology for 21st Century*, 165-170. Tsinghua Press, Beijing.
- Apley, D. W. and Shi, J. (1999). The GLRT for statistical process control of autocorrelated processes. *IIE Trans.* **31**, 1123-1134.
- Bagshaw, M. and Johnson, R. A. (1975). The influence of reference values and estimated variance on the ARL of CUSUM tests. *J. Roy. Statist. Soc. Ser. B* **37**, 413-420.
- Basseville, B. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, Englewood Cliffs, NJ.
- Capizzi, G. and Masarotto, G. (2003). An adaptive exponentially weighted moving average control chart. *Technometrics* **45**, 199-207.
- Chin, C. and Apley, D. W. (2007). An optimal filter design approach to statistical process control. *J. Quality Tech.* **39**, 93-117.
- Durrett, R. (1991). *Probability Theory and Examples*. Wadsworth, California.
- Han, D. and Tsung, F. (2004). A generalized EWMA control chart and its comparison with the optimal EWMA, CUSUM and GLR schemes. *Ann. Statist.* **32**, 316-339.
- Han, D. and Tsung, F. (2006). A reference-free Cuscore chart for dynamic mean change detection and a unified framework for charting performance comparison. *J. Amer. Statist. Assoc.* **101**, 368-386.
- Hawkins, D. M. and Olwell, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement*. Springer. New York.
- Jiang, W., Tsui, K. L. and Woodall, W. H. (2000). A new SPC monitoring method: the ARMA chart. *Technometrics* **42**, 399-410.

- Jiang, W., Wu, H., Tsung, F., Nair, V. and Tsui, K. L. (2002). Proportional integral derivative charts for process monitoring. *Technometrics* **44**, 205-214.
- Lucas, J. M. (1982). Combined Shewhart-CUSUM quality control scheme. *J. Quality Tech.* **14**, 51-59.
- Montgomery, D. C. (2001). *Introduction to Statistical Quality Control*. Wiley, New York.
- Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distribution. *Ann. Statist.* **14**, 1379-1387.
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100-115.
- Ritov, Y. (1990). Decision theoretic optimality of the cusum procedure. *Ann. Statist.* **18**, 1464-1469.
- Roberts, S. W. (1959). Control charts based on geometric moving averages. *Technometrics* **1**, 239-250.
- Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag, New York.
- Siegmund, D. and Venkatraman, E. S. (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Ann. Statist.* **23**, 255-271.
- Sparks, R. S. (2000). CUSUM charts for signalling varying location shifts. *J. Quality Tech.* **32**, 157-171.
- Srivastava, M. S. and Wu, Y. H. (1993). Comparison of EWMA, CUSUM and Shirayayev-Roberts procedures for detecting a shift in the mean. *Ann. Statist.* **21**, 645-670.
- Srivastava, M. S. and Wu, Y. H. (1997). Evaluation of optimum weights and average run lengths in EWMA control schemes. *Comm. Statist. Theory Methods* **26**, 1253-1267.
- Wu, Y. H. (1994). Design of control charts for detecting the change. In *Change-point Problems* (Edited by E. Carlstein, H. G. Muller and D. Siegmund), 330-345. IMS, Hayward, CA.

Department of Mathematics, Shanghai Jiao Tong University, Shanghai, China.

E-mail: donghan@sjtu.edu.cn

Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Kowloon, Hong Kong.

E-mail: season@ust.hk

Department of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China.

E-mail: ytli@sjtu.edu.cn

Department of Industrial Engineering, Tsinghua University, Beijing, China.

E-mail: kbwang@tsinghua.edu.cn

(Received March 2008; accepted April 2009)